

観光地検索システムの説明性向上のための既訪問スポットと検索結果の 対応付けの詳細化

潘 健太[†] 北山 大輔[†]

[†] 工学院大学大学院工学研究科情報学専攻 〒163-8677 東京都新宿区西新宿 1-24-2

E-mail: [†]tem18011@ns.kogakuin.ac.jp, ^{††}kitayama@cc.kogakuin.ac.jp

あらまし 近年、ユーザは Web 上の観光情報を活用して旅行計画を立てることが多くなっている。しかし、旅行は一般的に訪れたことがないスポットに行くことが多いため、観光情報を適切に理解することは困難である。そこで、訪問したことがある観光スポットの特徴を用いて、未訪問エリアの観光スポットを対応付けして、ユーザの未知なスポットに対する理解を支援することを提案する。本研究では、まず、観光スポットのユーザレビューを用いてレビューベクトルを作成する。次に、ユーザにとっての観光スポットの特徴を抽出するために、既訪問スポットレビューベクトルを用いて階層的クラスタリングを行うことによって、特徴クラスタを算出する。また、特徴クラスタに対応する検索結果スポットレビューを用いて類似度を計算し、既訪問スポットと検索結果スポットを関連付けを行い、関連性を説明するためのキーワードを抽出する。最後に、観光スポット間の対応関係を地図上に可視化する手法を用いて、プロトタイプシステムを構築し、既訪問スポットと検索結果スポットの説明性向上のための観光スポットの対応関係を評価する実験を行う。

キーワード 観光スポット, 理解支援, ユーザレビュー, 階層的クラスタリング, 可視化

1 はじめに

旅行先を決定するとき、旅行者は観光スポット検索サイトや観光情報に関連する書籍を見て観光スポットを選び、旅行計画を立てる。しかし、ユーザにとって訪問したいエリアは、訪問したことがなく不慣れであることが多い。そのため、エリア内に数多く存在する観光スポットから、自身の要求に合う観光スポットを見つけることは容易ではない。このとき、ユーザは観光スポット検索サイトのランキングやおすすめ情報を見て観光スポットを決めることが多くなると考えられる。ユーザの意思決定の材料の 1 つとして、Tripadvisor¹やじゃらん²などの観光スポット検索サイトがある。これらのサイトには特定の観光スポットを訪問したことのあるユーザがレビューを投稿し、観光スポットに関する豊富な情報が存在している。しかし、ユーザは検索エリアに関する事前知識がないため、どのスポットのレビューを読むべきか効率的に判断することは困難である。そのため、さまざまな観光スポットを効果的に理解するためには、ユーザが訪問した経験のあるスポットを使って検索結果スポットと比較することは効果的と考えられる。たとえば、日本に初めて訪れるフランス人旅行者に対し、検索結果スポットである東京の「表参道」をパリにおける「シャンゼリゼ通り」と表現すると理解しやすいであろう。この考え方は、ユーザが以前に経験した物事を現在の物事に適用する一種の類推である [1]。類推は創造的思考に貢献すると指摘されている [2]。類推に関する研究の多くは、ベースとなる学習データとターゲットとなる問

題が与えられ、物事の特徴を問題の特徴にマッピングして問題を解決するもの [3] である。

我々は、すでに既訪問スポットと検索結果スポットの対応関係を抽出する手法について取り組んでいる [4]。しかしながら、既訪問スポットそのものがユーザの関心を表すとは限らない。例えば、あるテーマパークに訪問したユーザでも、アトラクションに関心がある場合もあれば、キャラクターに関心がある場合も考えられる。そこで本稿では、既訪問スポットを用いてより詳細なユーザの関心を扱う手法に取り組む。本手法では、ユーザはすでに訪問した経験のあるスポットと未訪問エリアを入力する。そのとき、観光スポットのユーザレビューを用いて特徴ベクトルを作成する。次に、ユーザにとっての観光スポットの特徴を抽出するために、既訪問スポットのレビューベクトルを用いて階層的クラスタリングを行うことで、特徴クラスタを算出する。また、特徴クラスタに対応する検索結果スポットレビューを用いて類似度を計算し、既訪問スポットと検索結果スポットを関連付けを行い、関連性を説明するためのキーワードを抽出する。最後に、観光スポット間の対応関係を地図上に可視化する手法 [5] を用いて、プロトタイプシステムを構築し、既訪問スポットと検索結果スポットの説明性向上のための観光スポットの対応関係を評価する実験を行う。抽出したキーワードを提示することで、ユーザの検索結果スポットに対する理解の支援を目指す。図 1 は提案手法の概念図である。

本論文の構成は下記のとおりである。2 節では、関連研究について述べる。3 節では、提案手法の概要について述べる。4 節では、検索結果スポットと既訪問スポットの対応関係による地図上の可視化手法について述べる。5 節では、クラスタ分け

1 : <https://www.tripadvisor.com/>

2 : <https://www.jalan.net/kankou/>

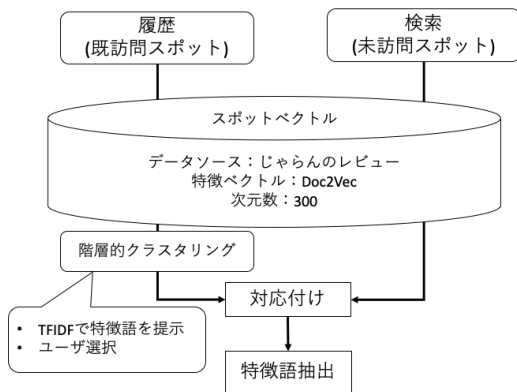


図 1 提案手法の概念図

に関する評価について述べる。最後に、6 節では、まとめと今後の課題について述べる。

2 関連研究

2.1 観光スポット検索と推薦システム

ユーザの体験履歴を利用した検索や推薦システムに関する研究は数多く発表されている。倉島ら [6] は、Flickr に投稿された写真のジオタグ情報を人々の旅行履歴として利用した旅行ルート推薦手法を提案した。この手法では、ユーザの現在地から行きやすい場所とユーザの興味に合致した場所に移動しやすいと仮定し、行動モデルを生成している。ユーザのジオタグ付き写真集合は、時間情報でソートすると個人の旅行履歴とみなすことができると考え、ジオタグ情報を利用してユーザの行動モデルを生成している。Kitamura ら [7] は、一般的な物体認識を用いて、過去の個人旅行写真から推定したユーザの旅行の嗜好に基づき観光地を推薦する方法を提案した。物体認識システムを用いて、写真で撮った被写体情報のキーワードを取得し、グラフ視覚化技術によってキーワードの共起を表現した。また、グラフの視覚化技術に基づいて旅行写真付きのグラフを視覚化するユーザインターフェイスを紹介した。

2.2 地理的情報の可視化と意味的な関係の可視化

本研究では、地理的情報の可視化と意味的な関係の可視化に取り組む。地理的な情報の可視化としては、抽出した情報を地図上にマッピングするのが一般的であり、本研究もそれに従っている。代表的な研究を以下に紹介する。櫻川ら [8] は、ソーシャルメディア上にアップロードされた写真のジオタグ情報と撮影時刻に基づいて写真の撮影者を分類する手法を提案した。分類された撮影者（観光者、在住者）ごとにホットスポットを可視化した。また、ジオタグ情報と撮影時刻以外にテキストタグを加えて、写真が撮影された地域で行われるイベントの穴場スポットを発見し、地図上に提示する手法を提案した [9]。

意味的な関係性の可視化としては、グラフモデル上でオブジェクト間のつながりを関連度で表現することが多い。上村ら [10] は、ユーザが投稿したタグ付き画像を用いたファッションスタイルの関係性の可視化手法を提案した。ファッションス

タイルは類似するタグを空間座標に固定することによって関係を表している。本研究では、スポット間の意味的な対応関係を可視化するために、検索結果スポットは地図上で実在する座標に固定され、配置の自由があるのは既訪問スポットのみという制約がある。

3 検索結果スポットと既訪問スポットの対応付け

我々は、ユーザの既訪問スポットに基づく検索結果スポットの対応付けの詳細化を提案する。まず、ユーザがすでに訪問した複数の観光スポットと訪問したい観光エリア情報を入力する。本手法では、観光スポットのユーザレビューを用いて既訪問スポットのレビューベクトルを生成する。検索結果スポットも同様にエリア内の各スポットのレビューベクトルを求める。次に、山田ら [12] が行ったユーザの嗜好抽出手法に基づいて、ユーザの観光スポット訪問履歴から各訪問スポットに付随する個々のレビューを全て取得し、階層的クラスタリングによってレビューをクラスタに分類する。各クラスタを正規化ジニ分散指標とスポット数で算出したスコアの上位 N 件のクラスタの特徴語をユーザに提示する。その後、未訪問エリア内のスポットレビューを用いて、ユーザが選択したクラスタに対して平均類似度によってクラスタの分類を行う。ユーザ選択クラスタに属する既訪問スポットと検索結果スポットのレビューベクトルの類似性によって、既訪問スポットを検索結果スポットと関連付ける。最後に、その関係性を説明するためのキーワードを抽出する。

3.1 スポットレビューを用いたレビューベクトル生成

本稿では、2016 年 9 月末までのじゃらんから得られた 1,481,838 件のレビューデータを使用し、分散表現 [11] を用いてレビューベクトルの作成する。本研究では、分散表現を計算するために Python のライブラリである gensim³ を利用する。学習方法として、Distributed Bag-of-Words を利用して、観光スポットのレビューを 300 次元のベクトルで表現する。既訪問スポットや検索結果スポットの各レビューベクトルは、形態素解析器である MeCab [13] に辞書「mecab-ipadic-NEologd」⁴ を用いて、分かち書き（原型）したレビューを利用して作成する。

3.2 既訪問スポットのクラスタ分類

本研究では、ユーザが訪れた観光スポットに、ユーザの観光スポットに対する嗜好が含まれていると仮定する。しかし、多くの観光スポットには、単一の要素ではなく複数の要素が存在する。そこで、山田ら [12] が行ったユーザの嗜好抽出手法に基づいて、ユーザが訪れた観光スポットに付随する全てのレビューをクラスタリングすることにより、レビューを観光スポットが持つ特徴ごとに分類する。山田らの手法を説明する。レビューの分類には階層的クラスタリングを利用する。クラスタ間の距離の測定には、レビューベクトルのコサイン距離による群平均

3 : <https://radimrehurek.com/gensim/models/doc2vec.html>

4 : <https://github.com/neologd/mecab-ipadic-neologd/>

法を用いる。群平均法とは、2つのクラスタに属する要素間の全ての組み合わせの距離を求め、その平均値をクラスタ間の距離とする手法である。クラスタ分けの閾値は、山田らが行った実験結果により0.65とする。

階層的クラスタリングによって分類されたクラスタに対し、どのクラスタがユーザの嗜好を表しているかをスコア付けする。本研究では、以下の条件を満たすクラスタをユーザの嗜好を表すクラスタと定義する。

- 各既訪問スポットのレビューを満遍なく含んでいる
- より多くのスポットのレビューで構成されている

本研究では、クラスタを構成する既訪問スポットごとのレビュー数の割合を算出するため、ジニ分散指標を用いる。ジニ分散指標とは各クラスの分散の総和を表す指標であり、クラスのばらつきが小さいと値は大きくなり、ばらつきが大きいほど値は小さくなる。既訪問スポット数を N 、クラスタ c に所属するレビューのうち、スポット s のレビューの割合を $P(s | c)$ とすると、ジニ分散指標は式1で表せる。

$$gini(c) = 1 - \sum_{s=1}^N P(s | c)^2 \quad (1)$$

また、各スポットのレビュー数とそのスポットの全レビュー数との割合を求めることにより、レビュー数の正規化を行ったものを正規化ジニ分散指標と呼称し、これを利用する。スポット s の総レビュー数を R_s とすると、正規化ジニ分散指標における $P(s | c)$ は式2で表せる。

$$P(s | c) = \frac{\frac{r_s}{R_s}}{\sum_{j=1}^N \left(\frac{r_j}{R_j} \right)} \quad (2)$$

ジニ分散指標は各クラスの分散の総和のみで値が決まる関係上、クラス数までは考慮されない。我々はより多くの既訪問スポットで構成されたクラスタが、ユーザの観光に対する嗜好を表現していると考えたため、クラスタの構成スポット数による重み付けを行う。そのとき、あるスポットに関するレビューが1つしかない場合でも、そのクラスタを構成するスポットとしてカウントしてしまう問題がある。そこで本研究では、クラスタを構成するあるスポットのレビュー数と、そのスポットの全レビュー数との割合が閾値以上のものを、クラスタを構成するスポットとする。本稿では、閾値を1%以上と決定した。

クラスタ c の嗜好の興味度 $pref(c)$ は、あるスポットに関するレビュー数の割合が閾値以上のスポット数を s_c とし、式3で示す。

$$pref(c) = gini(c) \times \log s_c \quad (3)$$

3.3 クラスタの特徴語抽出

山田ら[12]の手法によってユーザの嗜好を表すクラスタを抽出できる。本稿では、抽出したクラスタをユーザが任意に選択する必要がある。そのため、各クラスタの特徴語をユーザに提示する。各クラスタの特徴語として、スコア付されたクラスタに所属するレビューを用いて、TFIDF法で各単語のTFIDF値を求め、各クラスタの上位10件までの単語をユーザに提示す

る。各単語の特徴量を式4で定義する。

$$c_word(w, c) = \frac{tf_c(w)}{m_c} \times \log \frac{C}{df(w)} \quad (4)$$

ここで、クラスタ c における単語 w の出現回数を $tf_c(w)$ とし、クラスタ c の総単語数 m_c で割ることで正規化した TF 値を求める。全クラスタ数を C 、単語 w が出現するクラスタ数を $df(w)$ とし、 IDF 値を算出する。

3.4 対応するスポットの決定と特徴語抽出

求めたクラスタの興味度の上位 M 件のクラスタに所属するレビューと、未訪問エリア内のスポットの各レビューの類似度を計算する。未訪問エリア内のスポットのあるレビューと、クラスタに所属する各レビューの類似度の平均を、クラスタに対する検索結果スポットのレビューの類似度とする。検索結果スポットのレビューについて、類似度がもっとも大きい、かつ0.125以上となるクラスタにそのレビューを所属させる。このことにより、興味を表すクラスタに検索結果スポットのレビューを加える。検索結果スポットを説明するため、検索結果スポットと既訪問スポットの関連付けを行う。検索結果スポットと既訪問スポットの関連付けのため、検索結果スポットのレビュー数が5つ以下の場合、対応付けを行わない。まず、クラスタに所属するある既訪問スポットの各レビューベクトルと、ある検索結果スポットの各レビューベクトルを用いてスポット間の類似度を、群平均法により計算する。類似度計算には、コサイン尺度(式5)を用いる。

$$sim(s_f, s_u, c) = \frac{\sum_{r_f \in rev(s_f, c)} \sum_{r_u \in rev(s_u, c)} \cos(r_f, r_u)}{|rev(s_f, c)| \times |rev(s_u, c)|} \quad (5)$$

s_f を既訪問エリア内のあるスポットとし、 s_u を未訪問エリア内のあるスポットとする。 rev は引数に与えられたスポットのクラスタ C に属するレビューを返す関数である。 r_f はクラスタに所属し、かつ既訪問スポット s_f のレビューベクトルである。 r_u はクラスタに所属し、かつ検索結果スポット s_u のレビューベクトルである。

スポットの対応関係を示すだけでは、どのような点で対応するのかを理解するのは難しい。そこで、検索結果スポットと既訪問スポットの関係性を表すキーワードをユーザに提示する。すべてのレビューを3.1節と同様に形態素解析器 MeCabによって単語に分割する。ただし、助詞、助動詞、連体詞、記号、ストップワードを削除している。

続いて、キーワード抽出手順について説明する。まず、TFIDF法を使って対象となる既訪問スポットと検索結果スポットの特徴語のTFIDF値を求める。既訪問スポット集合を S_f 、検索結果スポット集合を S_u 、それぞれの要素を s_f, s_u とする。IDF値を算出する集合は s_f の語については S_f, s_u の語については S_u を用いる。

2つのスポットに共通する特徴語のTFIDF値の調和平均を用いて、対応付けしたスポットの説明可能なキーワードを抽出する。各単語のスコアを式6によって定義する。

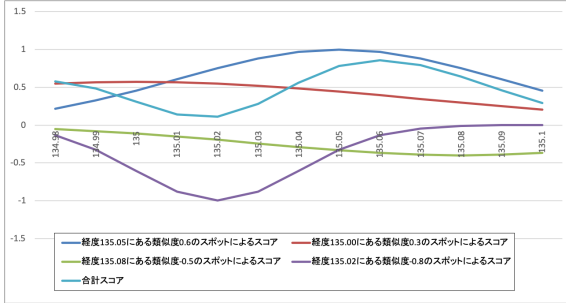


図 2 既訪問スポットの経度座標計算の例

$$score(t, s_f, s_u) = \frac{2 \times tfidf(t, s_f) \times tfidf(t, s_u)}{tfidf(t, s_f) + tfidf(t, s_u)} \quad (6)$$

$tfidf(t, s_f)$ と $tfidf(t, s_u)$ は同じ単語に関する既訪問スポット s_f における TFIDF 値と検索結果スポット s_u における TFIDF 値を示している。単語スコアの上位 N 個の単語を説明情報としてユーザに提示する。

4 スポット間の対応関係を用いた可視化

3 節で求めたある検索結果スポットに対して、入力した既訪問スポットがどのような関係にあるのか地図上で、一目で把握できるシステムが望ましい。そのため、我々は観光スポット間の対応関係を地図上に可視化する手法 [5] を提案した。この手法では、ユーザは地図上で検索結果スポットを検索しているものとし、まず、地図上の実際にスポットが存在する座標に検索結果スポットを配置する。次に、既訪問スポットの位置を決定する。これは検索結果スポットとの対応関係によって変化し、検索結果スポットとの類似度が高いと近く、類似度が低いと遠くなるように配置することで、意味的な位置関係を表現する。既訪問スポットの座標は、式 7 によって求める。

$$Score(s_f, T) = \sum_{s_u \in S_u} w(s_u, s_f) \times P_{s_u, s_f}(d(s_u, T)) \quad (7)$$

$w(s_u, s_f)$ は、ある既訪問スポット s_f と検索結果スポット s_u の類似度が正であれば 1.0、負であれば -0.5 を返す関数である。 T は、候補の座標である。 $d(s_u, T)$ は、ある座標 T から検索結果スポット s_u のユークリッド距離である。また、 P_{s_u, s_f} は、平均 μ が 0、標準偏差 σ が $(1 - |\cos(\text{vec}(s_u), \text{vec}(s_f))|) \times \alpha$ である正規分布を表す。 α は本稿では 1 とする。 vec は検索結果スポットや既訪問スポットの特徴ベクトルを返す関数である。 $Score(s_f, T)$ がもっとも大きな座標 T を既訪問スポットの座標とする。図 2 は既訪問スポットの経度座標計算の例である。縦軸はスコアの値であり、横軸は経度である。4 つの検索結果スポットを用いて、合計スコア $Score(s_f, T)$ を計算している。その結果、135.06 の座標がもっとも高くなる。緯度についても同様に行う。本稿では、近似として対象エリア内でグリッド上に 289 個の T を作り、それを用いた。

表 1 ユーザ A の既訪問スポット履歴

既訪問スポット	想定した嗜好	レビュー数
成田山新勝寺	初詣	787
川崎大師	初詣	241
東京スカイツリー	タワー	237
東京タワー大展望台	タワー	1875
海遊館	生物	2874
沖縄美ら海水族館	生物	3028
新宿御苑	その他	1147

表 2 ユーザ B の既訪問スポット履歴

既訪問スポット	想定した嗜好	レビュー数
富士急ハイランド	遊園地	1328
ユニバーサル・スタジオ・ジャパン	遊園地	760
高千穂神社	神社	135
天岩戸神社	神社	228
阿蘇山	山	444
富士山	山	761
新宿御苑	その他	1147

5 クラスタ分けに関する予備実験

5.1 実験内容

ユーザが訪問したことのある観光スポットから、ユーザの観光スポットに対する嗜好をクラスタリングする実験を行う。まず、既訪問スポットに対する嗜好を 3 つ想定したデータセットを、我々が仮想的に 2 人分用意する。このデータセットを用いて既訪問スポットのレビューを取得し、階層的クラスタリングによってクラスタを作成しスコア付けを行う。そのあと、スコア上位のクラスタを構成するレビューを分析し、想定した嗜好を含むクラスタが作成されているかどうかを確認する。

本実験で利用するデータセットの内容を以下に示す。ユーザ A のデータセットを表 1、ユーザ B のデータセットを表 2 のように用意した。ユーザ A は「初詣」、「タワー」、「生物」を観光スポットに対する嗜好と仮定したユーザを想定している。ユーザ B も同様に、「遊園地」、「神社」、「山」を観光スポットに対する嗜好とするユーザを想定した。また、ユーザの観光スポット訪問履歴には様々な嗜好が混じっていることから、「その他」として異なるジャンルの観光スポットをそれぞれ 1 つずつ含めている。

5.2 実験結果と考察

表 3 と表 4 は正規化ジニ分散指標とスポット数による重み付けを用いたスコア付けの結果となっている。階層的クラスタリングによるクラスタ分類がうまく出来ているかを判断するため、スコアが大きい上位 5 件を取り上げて考察する。それぞれのクラスタ内の TFIDF 法で計算した TFIDF 値の上位の 10 件の単語を抽出して、クラスタがうまく作成されているかどうかを判断する要素として利用する。

ユーザ A のクラスタ分類の実験結果 (表 3) から、195 番のクラスタでは「初詣」に関する特徴語、134 番のクラスタでは

表 3 ユーザ A のクラスタ分け結果

クラスタ	スコア	クラスタ内の特徴語上位 10 件
151	0.57	水槽, ジンベイザメ, 水族館, 魚, ジンベエザメ, 迫力, サメ, 泳ぐ, 沖縄, イルカショー
134	0.56	東京タワー, 展望台, 夜景, スカイツリー, 東京, 景色, タワー, シンボル, 展望, 見える
195	0.14	成田, 成田山, うなぎ, お寺, 新勝寺, 成田山新勝寺, 初詣, 参道, 節分, 成田駅
155	0.07	記入, 館内, エスカレーター, 残念, 従業員, 悪さ, 対応, お天気, コーナー, 水族館
186	0.06	川崎大師, 渡す, お祓い, バッグ, 仲見世通り, 御籤, 一服, なんだか, APEC, すれ違う

表 4 ユーザ B のクラスタ分け結果

クラスタ	スコア	クラスタ内の特徴語上位 10 件
131	0.38	絶叫, アトラクション, 乗れる, トーマスランド, 乗り物, 乗る, ええじゃないか, ジェットコースター, ドドンパ, 遊園地
80	0.23	富士山, 火口, 合, 登る, 五合, 山頂, ガス, 阿蘇, 阿蘇山, 見える
56	0.14	天岩戸, 宮, 高千穂峡, 高千穂, 神社, 西本, 杉, 天安河原, 社務所, 天岩戸神社
53	0.08	参拝, 詣る, 参拝者, あまる, うんざり, ひとつだけ, 事情, 仏閣, 巫女, 手水舎
54	0.07	神話, 一回り, 学べる, 宮崎市内, 気象観測, 津, 美美, 駅そば, かけて, 説明

「タワー」に関する特徴語, 151 番のクラスタでは「生物」に関する特徴語が抽出されており, 我々が仮定した嗜好に合うクラスタ分類ができています。同様に, ユーザ B のクラスタ分けの実験結果 (表 4) から, 131 番のクラスタでは「遊園地」に関する特徴語, 56 番のクラスタでは「神社」に関する特徴語, 80 番のクラスタでは「山」に関する特徴語が抽出されており, 我々が仮定した嗜好に合うクラスタ分類ができています。この結果から, クラスタの構成スポット数による重み付けが, ユーザの観光スポットに対する嗜好の抽出に有用であることがわかった。また, クラスタのスコアを見ると, スコア値が 0.1 以上のときには, クラスタの特徴語がうまく抽出されているに対し, スコア値が 0.1 より小さいときには, 抽出された単語のばらつきが大きくなっている。そのため, ユーザのに提示するクラスタはスコア値 0.1 以上の上位 3 件とする。ユーザがスポット検索を行う際に, 自分の嗜好に合うクラスタを選択することによって, より検索結果の対応付けを詳細にすることが可能だと考えられる。

6 説明キーワードに関する実験

6.1 実験内容

ユーザが訪問したことのある観光スポットから, 検索結果スポットの説明キーワードに対する実験を行う。抽出された説明キーワードを評価するために, 以下の 2 つの手法を使って比較する。

- 提案手法: レビューを階層的クラスタリングにより分類する

- 既存手法: レビューを分類しない (スポット毎を 1 文書)

既存手法は, 我々は, すでに既訪問スポットと検索結果スポットの対応関係を抽出する手法 [4] を用いる。既存手法は, 観光スポットのユーザレビューを用いて特徴ベクトルを生成する。さらに, ユーザの観光地の位置付けを抽出するために, あるスポットに対し他の訪れたスポットと比較した相対的特徴ベクトルを算出する。同様に, 検索結果スポットもエリア内の各スポットの特徴ベクトルを求める。また, 未訪問エリアの各観

表 5 既訪問スポット集合と検索結果スポット集合

既訪問スポット	検索結果スポット
東京ドイツ村	屋島寺
富士急ハイランド	栗林公園
国営ひたち海浜公園	丸亀城
アクアパーク品川	金刀比羅宮
ハウステンボス	総本山善通寺
長崎バイオパーク	日本ドルフィンセンター
川越氷川神社	エンジェルロード
千里浜なぎさドライブウェイ	ニューレオマワールド
	プレイパークゴールドタワー
	国営讃岐まんのう公園

光スポットの相対的特徴ベクトルを, そのエリアの他の観光スポットと比較して計算する。次に, 相対的特徴ベクトルの類似度によって既訪問スポットを検索結果スポットと対応付ける。最後に, その関係性を説明するためのキーワードを抽出する。

本実験は, まず, 5 個から 9 個の自分がすでに訪れたことのある観光スポットを入力する。次に, 旅行など行ったことがなく, これから訪れたい都道府県やエリアを入力する。最後に, 検索結果スポットに対応する既訪問スポットと説明キーワードを提示し, 2 つの手法の比較を行う。

表 5 は, 既訪問スポット集合と検索結果スポット集合の検索例を示している。検索結果スポットは香川県からランダムに選んだ 10 つのスポットである。

6.2 実験結果と考察

表 6 は, 3.2 節で分類したスコアの上位 3 件, かつ 0.125 以上のクラスタを示している。また, 3.3 節で提案した方法を用いて抽出したクラスタの特徴語も示している。表 7 は, 提案手法を用いて, クラスタ 309 を選ぶとき, 対応付けられたスポットと上位 15 個の説明可能なキーワードの一例を示している。表 8 は, 既存手法を用いて, 対応付けられたスポットと上位 15 個の説明可能なキーワードの一例を示している。

クラスタ 309 の特徴語を見ると, 風景, 時期や楽しいといった要素を含んでいる。よって, 関係を説明するキーワードもそ

表 6 分類した上位 3 件のクラスタ

クラスタ	クラスタ内の特徴語上位 10 件
309	イルミネーション, ネモフィラ, ハウステンボス, チューリップ, KOKIA, 花, きれい, 時期, 光, 楽しめる
293	イルカ, 水族館, イルカショー, ショー, 品川駅, アシカ, 動物, 水槽, カピバラ, リスザル
295	芝生, 車, フリーパス, 転がる, パターゴルフ, 芝桜, 移動, 子供, 遊具, 乗れる

表 7 提案手法の検索結果の一例

既訪問スポット	関係を説明するキーワード	検索結果スポット
富士急ハイランド	時期, いい, 泊まる, レンタル, また, 楽しめる, 乗る, ぜひ, 有名, 行ける, 人気, 素敵, 来る, 天気, 連続	ニューレオマワールド
東京ドイツ村	訪れる, 気持ち, 知る, また, 場所, 楽しめる, よい, 食べる, 絶対, レンタル, 世代, 雰囲気, いい, 観光, ぜひ	国営讃岐まんのう公園

表 8 既存手法の検索結果の一例

既訪問スポット	関係を説明するキーワード	検索結果スポット
富士急ハイランド	自転車, 美術館, アート, レンタサイクル, 作品, プロジェクト, バス, ミュージアム, おしゃれ, 海外, 建築, 巡る, 日常, 泊まる, まわる	ニューレオマワールド
東京ドイツ村	映画館, 映画, 展示, 演出, 魚の, 臭い, 入場料, 内容, 使用, 泳ぐ, 観る, 金魚, 展示物, 見どころ, 建物	国営讃岐まんのう公園

れを示すものがふさわしい。表 7 は、クラスタ 309 に基づき対応付けられたスポットであり、「ニューレオマワールド」と「富士急ハイランド」の関係を説明するキーワードには、時期, 楽しめる, 素敵が適切に提示されている。また、「国営讃岐まんのう公園」と「東京ドイツ村」の関係を説明するキーワードには、楽しめる, 世代, 雰囲気が適切に提示されている。

一方表 8 は、既存手法による関係を説明するキーワードの抽出結果であるが、「ニューレオマワールド」と「富士急ハイランド」は、自転車, 美術館やバスなど多様なキーワードが提示されている。また、「国営讃岐まんのう公園」と「東京ドイツ村」も同様に、映画館, 展示や泳ぐなど多様なキーワードが提示されている。

提案手法と既存手法の結果から、スポットの検索目的によって異なる手法を用いることが最適な検索が可能と考えられる。提案手法は、レビューに対しクラスタリングを行うことによって、特定の要素に関するスポットを対応付け、そしてキーワード抽出することが可能と考えられる。既存手法では、スポットの全体的な特徴を表すことができる。よって、多様な情報を提示することが可能で、意外性のあるスポットを検索することが可能と考えられる。

7 まとめと今後の課題

本研究では、ユーザが行きたい観光スポットが決まっていな場合に、ユーザの事前知識が不足しているため、観光検索サイトを使用してランキング、おすすめ情報やカテゴリなどを基に検索された観光スポットに対する理解が困難であることに着目した。検索結果スポットに対する理解を支援するために、検索結果スポットをユーザがすでに訪れたことがある既訪問スポットと比較することによって、理解を支援する説明手法を提案した。具体的に、ユーザの既訪問スポットから各スポットのレビューを全て取得し、そのレビューを用いて階層的クラ

スタリングを行った。次に、得られたクラスタに対しユーザの嗜好を表しているかどうかのスコア付けを行い、スコア上位のクラスタの特徴語をユーザに提示した。本稿ではシステムの実装に向けて、既訪問スポットからユーザの嗜好推定が可能かどうかについて実験を行った。その結果、スコア上位のクラスタはユーザの嗜好を表していることが判明した。また、関係を説明するキーワードに関する実験を行った結果、提案手法の方は、レビューを階層的クラスタリングを行うことによって、特定の要素に関するスポットを検索することが可能と判明した。既存手法は、スポット全体的な特徴を表すことができるため、多様な情報と意外性のあるスポットを検索することが可能と判明した。

今後の課題は、システムのインタフェースの作成およびユーザの観光スポット検索を詳細化できるか明らかにすることである。詳細化できるかを明らかにするために、提案手法と既存手法との評価実験を行う必要がある。

謝 辞

本研究の一部は、2019 年度科研費基盤研究 (B)(課題番号: 19H04118) によるものです。ここに記して謝意を表すものとします。

文 献

- [1] D. Gentner, “Structure-Mapping: A Theoretical Framework for Analogy”, Cognitive Science, Vol.7, pp.155–170, 1983
- [2] K. J. Holyoak and P. Thagard, “Mental Leaps: Analogy in Creative Thought, MIT Press”, Journal of Japanese Society for Artificial Intelligence, Vol.11, No.3, pp.489, 1996
- [3] M. L. Gick and K. J. Holyoak, “Analogical Problem Solving”, Cognitive Psychology, Vol.12, pp.306–355, 1980
- [4] 潘 健太, 北山 大輔: ユーザの既訪問スポットの位置付けに基づく検索結果スポットの説明手法, 第 11 回データ工学と情報マネジメントに関するフォーラム (DEIM Forum 2019), H7-2, 2019
- [5] 潘 健太, 北山 大輔: 地図上における検索結果スポットの説明性

向上のための観光スポットの対応関係可視化手法, 観光情報学会
第 20 回研究会, pp.1-4, 2019

- [6] T. Kurashima, T. Iwata, G. Irie and K. Fujimura., “Travel route recommendation using geotags in photo sharing sites”, CIKM ’10 Proceedings of the 19th ACM international conference on Information and knowledge management, pp.579-588, 2010
- [7] Kitamura, R. and Itoh, T.: Tourist Spot Recommendation Applying Generic Object Recognition with Travel Photos, In 2018 22nd International Conference Information Visualisation, pp.1-5, 2018
- [8] 櫻川 直洋, 廣田 雅春, 石川 博, 横山 昌平: ジオタグ付き写真の撮影者を在住者と観光者に分類することによるホットスポットの発見, 第 7 回データ工学と情報マネジメントに関するフォーラム (DEIM Forum 2015), F6-3, 2015
- [9] 櫻川 直洋, 廣田 雅春, 石川 博, 横山 昌平: ジオタグ付き写真を用いたイベントとその穴場スポットの発見, 第 8 回データ工学と情報マネジメントに関するフォーラム (DEIM Forum 2016), H5-3, 2016
- [10] 上村 幸汰, 桂井 麻里衣, 真木 勇人, 後藤 亮介: タグ付き画像を用いたファッションスタイルの関係性の可視化, 第 11 回データ工学と情報マネジメントに関するフォーラム (DEIM Forum 2019), E8-2, 2019
- [11] Quoc V. Le and Tomas Mikolov, “Distributed representations of sentences and documents”, In Proceedings of the 31th International Conference on Machine Learning, ICML 2014, pp. 1188–1196, 2014
- [12] 山田 祥輝, 北山 大輔: ユーザの嗜好に基づく観光スポット説明文の個人化手法, 第 11 回データ工学と情報マネジメントに関するフォーラム (DEIM Forum 2019), E3-1, 2019
- [13] T. Kudo, K. Yamamoto and Y. Matsumoto, “Applying Conditional Random Fields to Japanese Morphological Analysis”, Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing (EMNLP-2004), pp.230-237, 2004